

J005 Rec'd PCT/PTO 23 SEP 2005

Description

Speaker-dependent voice recognition method and voice recognition system

5

The invention relates to a speaker-dependent voice recognition method with a voice recognition system, in which voice utterances of a user are trained and commands are assigned to the trained voice utterances, 10 and to a voice recognition system for carrying out the method.

According to the prior art, such a method is divided into a voice recognition mode and a training mode. In 15 the voice recognition mode, voice utterances of the user are detected whereupon a command assigned to the voice utterance is found in a database if the voice utterance exhibits sufficient correspondence with a voice utterance which belongs to the command and was recorded and stored at an earlier time. In the voice 20 recognition mode, a new assignment between a new voice utterance and a new command is not possible. Instead, these processes take place in the training mode in which the user utters voice utterances and assigns a 25 command to each individual voice utterance after it has been recorded. The assignment obtained is stored in the database. Assigned commands can be, for example, dialing processes for subscribers to a communication network or voice control commands.

30

The method according to the prior art has the disadvantage that the training of new commands is complicated inasmuch as it is necessary for the user to actively switch from the voice recognition mode to the 35 training mode every time. This also has a negative effect on the market acceptance of voice recognition systems.

On the basis of this, the invention is based on the object of specifying a speaker-dependent voice recognition method and a voice recognition system for this in which new commands can be trained in a time-saving manner.

This object is achieved with regard to the method of the type initially mentioned in that on non-recognition of a voice utterance, the voice recognition system provides the user with the opportunity to immediately assign the voice utterance to a new command.

When carrying out the method, a voice recognition system is always in the voice recognition mode, but the option is available immediately to perform a new command assignment on non-recognition of a voice utterance. In this manner, the training of new commands is integrated into the voice recognition itself and can take place when a voice utterance has not been recognized. If, for example, the user happens to be in the situation where he wishes to train a new command for a voice recognition system, it is sufficient to articulate a voice utterance hitherto not used, whereafter the voice recognition system finds a non-recognition of the new voice utterance and then offers the option of assigning the voice utterance to a new command. After the assignment has been performed, the command can be executed immediately.

In a preferred embodiment, on non-recognition of the voice utterance by the voice recognition system the user can optionally either repeat the voice utterance or assign a new command to the voice utterance. This embodiment takes into account that a voice utterance can be just outside the range of similarity to a voice utterance to which a desired command has already been assigned. In this case, it is not intended to assign a new voice utterance to a new command. Instead, this voice utterance must be repeated in order to be linked

to the already trained command.

Having regard to an initial state of a voice recognition system, it is preferably provided for the
5 method that in the case when no command has yet been assigned to any voice utterance, the voice recognition system, after having been activated, immediately offers the training of a new command. This happens automatically when the voice recognition system
10 naturally does not recognize the first voice utterance and offers the option of training a new command.

In another embodiment, it can be provided that, on non-recognition of a voice utterance for a command already
15 trained by the voice recognition system, the user can select the command and assign the voice utterance to this command. This refers to the case where a "poor" version of the voice utterance is present in a database which contains the assignments between voice utterances
20 and associated trained commands, so that a voice recognition frequently fails. It is possible in this case to assign a new voice utterance to the command already trained.

25 For recognition of a voice utterance, a voice pattern is preferably generated which is assigned to the voice utterance. Such voice patterns, which are based on an extraction of essential voice features of the voice utterance, are then also used in the database which in
30 this case contains an assignment between voice patterns and trained commands. After having been recorded, each voice utterance is converted into a voice pattern which is then processed further, for example for the decision whether it is recognizable or not, i.e. is already
35 present within a range of similarity of a voice pattern in the database.

In this connection, it is regarded as being preferable to check before a command is assigned to a voice

utterance whether a voice utterance is similar to previously stored voice utterances before a command is assigned to a voice utterance. This prevents confusion among different commands from occurring during a voice
5 recognition because the associated voice utterances are in each case too similar to one another. For this purpose, a permissible range of similarity can be defined, for example by using the extraction features for a voice pattern.

10

The abovementioned object is achieved with regard to a voice recognition system by a voice recognition system for a speaker-dependent recognition of voice comprising a voice recording device for recording a voice
15 utterance of a user of the voice recognition system, a search engine which is designed for accessing a database which contains an assignment between voice utterances and commands in order to find a command assigned to the voice utterance, a conversion device
20 for converting the command found due to the voice utterance, the voice recognition system being designed in such a manner that on non-recognition of the voice utterance, the voice recognition system provides the user with the opportunity to immediately assign the
25 voice utterance to a new command.

Such a voice recognition system allows the method described above to be carried out and, compared with known voice recognition systems, is distinguished by
30 the fact that the training of new commands is made possible in a voice recognition mode.

The voice recording device is preferably connected to a memory in which the voice utterance is temporarily
35 stored and which is connected to the database for reading the voice utterance into the database. This is not the case in known voice recognition systems because in these, the database is directly accessed for a training mode whereas in a voice recognition mode,

although a voice utterance is temporarily stored for the operation of the search engine, the memory then used is not designed/linked for reading a voice utterance into the database.

5

Preferably, a feature extraction device for generating a voice pattern from the voice utterance is provided between the voice recording device and the memory and the voice pattern replaces the voice utterance.

10

Further advantages and features of the voice recognition system have already been explained above by means of the description of the speaker-dependent voice recognition method.

15

In the text which follows, an illustrated embodiment of the invention will be explained in greater detail with reference to the drawing. The only figure shows a flow chart of a speaker-dependent voice recognition method.

20

A speaker-dependent voice recognition method by means of a voice recognition system will now be explained with reference to figure 1. After a start of the voice recognition system, which is implemented, for example, 25 as a computer system with a display device, a suitable user interface which also contains an activation for a recording of a voice utterance ("push-to-talk" activation) is first displayed to the user. In a first method step 1, a voice utterance of the user/speaker is 30 recorded with the aid of a suitable voice recording device. In a second step 2, a voice pattern of the voice utterance is generated by means of a feature extraction device, a voice pattern being defined by a combination of extracted characteristic voice features. 35 The voice pattern is temporarily stored in a memory.

In a third step 3 a search engine is used to interrogate whether the voice pattern generated is contained in a database which contains assignments

between voice patterns and commands. This database is provided with contents in a training mode of the voice recognition system, the training mode being integrated into the process of a voice recognition. If the voice
5 pattern is recognized as already present in the database and the associated command is found, the command is executed in a fourth step, after which the operating process of the voice recognition system is ended. The sequence from step 1 to step 4 is automatic
10 in the present illustrated embodiment.

If the voice pattern generated is not recognized in the third step 3, the user receives the option of assigning a new command to the unrecognized voice pattern or the
15 unrecognized voice utterance, respectively, via the user interface of the computer system. This takes place in a fifth step 5 of the method. At this point, the voice recognition system is switched into a training mode if the assignment of a new command is desired or
20 automatically performed. As an alternative to the fifth step 5, the user can also trigger a new voice utterance recording with the aid of the user interface so that the process returns to the first step 1 in order to repeat the voice utterance.

25 If the assignment of a new command to the unrecognized voice pattern is selected, a voice utterance corresponding to the unrecognized voice utterance from the first step is recorded in a sixth step 6. Following
30 this, a voice pattern is generated from the voice utterance recorded in the sixth step 6 in a seventh step 7, in the same manner as in the second step 2 explained above.

35 In an eighth step 8, a similarity check between the new voice pattern from the seventh step 7 and the voice pattern from the second step 2 is performed. If a desired degree of correspondence between the two voice patterns is not obtained, the method begins again until

a satisfactory result for the similarity of the voice patterns generated in the second step 2 and the seventh step 7 is obtained. During this process, the third step 3 and the fifth step 5 can be skipped.

5

In the eighth step 8, a similarity check can also be performed to see whether the voice pattern of the newly recorded voice utterance is sufficiently distinct compared with the voice patterns already present in the database. If not, the user can be requested to use a different voice utterance for assignment for a new command. The method recommences with this new voice utterance.

15 Following this, a command is assigned to the voice pattern generated in the second step 2 in a ninth step 9 by a suitable selection of the user with the aid of the user interface of the voice recognition system. For this purpose, the voice pattern is read from the memory 20 in which it was temporarily stored in the second step 2, suitably combined with the voice pattern generated in step 7, e.g. by averaging individual characteristics of both voice patterns and written into the database together with the new command.

25

In a final step 10, the newly assigned command is executed after which the voice recognition process with integrated training mode is concluded.

30 It must be emphasized that the execution of a command taking place in the fourth and last step takes place with the aid of a conversion device for converting the command. The command can be, for example the dialing of a telephone number in a communication network or a 35 voice command by means of which devices connected to a network are controlled.

Naturally, in a simplified embodiment of the method, the performance of the preceding steps 6 to 8 can be

omitted when a command is assigned according to the ninth step 9. In this manner, a command is assigned immediately following the interrogation from the fifth step 5. It is also possible to dispense with the 5 immediate execution of the newly trained command (tenth step) during the performance of the method.